

CHAPTER THREE: METHODOLOGY

3.1 Introduction

The primary objective of this chapter on methodology is to provide a clear and accurate description of the instruments and methodology employed in carrying out this research. The study was carried out over a period of two months at Taylor's College Subang Jaya Sdn. Bhd (TCSJ). The aim of this research was to examine the effects of two popular methods of scoring (holistic and analytic) and identify the relationship(s) (if any) between the method of scoring and improvement in the writing performance of students taking short term language courses.

In order to ensure validity of the research design and analysis, a methodological triangulation research design was adopted. According to Neuman and Kreuger (2003, p.137), triangulation means "...to look at something from several angles than to look at it in only one way." Therefore, triangulation is any attempt to investigate a phenomenon using more than one method. It is used to counteract threats to validity that each experimental method contains. This study incorporated the elicitation of data from three sources - students, scorers and teachers, through the use of three methods - questionnaires, interviews, pre-tests and post-tests.

3.2 The International English Language Testing System (IELTS) Exam

The IELTS exam was formally released for use in November 1989. The test is jointly assessed and run by the Cambridge ESOL, the British Council and IDP Education Australia. It is now available in Britain, throughout Australia, in all British Council offices around the world, in IDP offices established in many Asian and Pacific countries and in other countries where trained examiners are available. Most academic institutions in Australian, British, Canadian and New Zealand accept the IELTS. Recently, an increasing number of academic institutions in the United States have begun to accept the IELTS for admissions purposes too. IELTS is also recognised and accepted by many professional organizations, including the New Zealand Immigration Service, the Australian Department of Immigration and Multicultural Affairs, the Australian Medical Council, the United Kingdom General Medical Council and the United Kingdom Ministry of Defense.

IELTS is a comprehensive test of English language proficiency designed to assess the ability of non-native speakers of English who intend to study or train in the medium of English. IELTS assesses the complete range of English language skills – listening, speaking, reading and writing, which are commonly encountered by students who intend to pursue their studies or undertake training in the English medium. The test, therefore, ascertains whether students have a sufficient

proficiency level in English to undertake training and academic programmes without their study being unduly inhibited by their English proficiency.

There are two versions of the IELTS exam – the Academic Module, which is for students seeking entry to institutions of higher education and the General Training Module, for students seeking entry to secondary schools, vocational training courses or to undertake work experience. The General Training modules are also used for immigration purposes.

The four-macro skills (listening, speaking, reading and writing) are emphasized in both these versions. Every test candidate is required to take a test for each of these skills. The listening and speaking modules are the same in both, the Academic Module and the General Training Module. However, candidates may choose between the Academic or General Training versions of the reading and writing sections of a test.

In the listening component, a student is required to listen to and comprehend spoken language in both lecture formats as well as formal and informal conversational styles. The recorded texts include a variety of short lectures and conversations and feature a variety of English accents and dialects.

The speaking test takes the form of a face - to - face interview. Candidates are assessed on their ability to use the English language to answer short questions, speak at length on a familiar topic and to interact with the examiner.

The reading module expects students to understand academic or training language. In the academic module, the reading texts are taken from books, magazines, journals and newspapers. All articles/texts are for a non- specialist audience. The general training module, on the other hand, consists of material/text that candidates would encounter on a daily basis in an English speaking country. The materials are usually sourced from instruction manuals, advertisements or newspapers.

Lastly, the writing section consists of two parts. For the first task, the candidates are to write a report of around 150 words based on the material found in a table, diagram, or chart. They are assessed on their ability to describe and explain data. Task two requires the candidate to write an essay of about 250 words in response to an opinion or problem. The candidates are required to be able to discuss issues, construct arguments and use appropriate tone and register.

The IELTS exam results provide a profile of a candidate's ability to use English. The test scores are given in the form of band scores. A band score from 0 to 9 is given for each component. The individual module scores are then averaged and rounded to produce an overall band score. The overall band scores and listening

and reading band scores are reported in whole and half bands (5, 6.5, 4, 7.5 etc.). The writing and speaking band scores are reported in whole bands only (4, 7, 5, 3, 7 etc.).

3.2.1 The Taylor's College Subang Jaya IELTS Preparatory Course

At Taylor's College Subang Jaya (TCSJ) students take the IELTS exam because they wish to pursue a degree in English speaking countries. Therefore, all students take the 'Academic' version of the test. The pre – university students at TCSJ need to obtain a minimum overall band of 6 to 6.5 in order to pursue their tertiary education. Without this qualification, it would not be possible for them to obtain admission into a foreign university especially in countries such as Australia, United Kingdom and New Zealand. The Language Centre of TCSJ offers the IELTS preparatory courses to help students prepare for the IELTS exam.

There are two types of preparatory courses - one is a 32-hour course and the other is a 16-hour preparation course. The duration for each course is 2 months and one month, respectively. Students have four hours of classes every week. Classes are either held on two weekdays from 3.00 to 5.30pm or on Saturdays from 8.00 to 12.30pm (with an interval of half an hour).

3.2.2 The Course Content

The four-macro skills are covered in general in the IELTS course (listening, speaking, reading and writing). Students are given several practice exercises and sample tests in order to make them familiar with the demands of the IELTS exam. The 16-hour preparatory course was chosen for the purposes of this study as it was the shorter of the two courses and the researcher's objective was to evaluate the effects of the holistic and analytic methods of scoring, on writing performance in 'short term' language preparatory courses. Thus, the focus of this study was only on the writing module of the IELTS exam.

The writing module of the IELTS exam comprises two tasks. In Task 1, the student is expected to write a report outlining the key features of a given table, diagram or chart. The required length is 150 words and the student has 20 minutes to complete the task.

However, for the purposes of this study, only Task 2 was used for the investigation. Task 2 requires students to write an essay of about 250 words using the academic writing format. Hence, it takes into consideration most of the areas that are pertinent to essay writing, namely, content, language, vocabulary, language and mechanics. A candidate has 40 minutes to complete this task. In Task 2, the student will have to consider an opinion in relation to evidence, or weigh the pros and cons of an argument and then write the essay. The question

may also require the student to discuss various aspects of a problem and then outline ways of solving it.

Out of the 16 hours of class time, six hours are set aside for the teaching of Writing Task 2. Students face the most problems with this task. The teaching of the writing skills in this task includes the current traditional rhetoric method of essay writing for three types of essays - discursive, cause-effect and argumentative. Students are exposed to the organization format of each type of essay and then made to write it. They are given six writing practice exercises over the four- week duration. All the writing practice exercises are done out of class (as homework) since the six hours allocated for class time is only sufficient for classroom instruction and discussions.

3.3 Subjects

The subjects of this study consists of students enrolled in the 16-hour IELTS preparatory course with the English Language Centre of TCSJ, scorers who volunteered to grade the student essays using two methods of scoring and lecturers of TCSJ who have been involved in conducting the IELTS intensive preparation courses. Each of these subjects had to fulfill a set of criteria in order to be included in the study.

Sampling is the process of selecting units from a population of interest so that by studying the sample we may fairly generalize our results back to the population from which they were chosen (Neuman and Kreuger, 2003). Fraenkel and Wallen (2000) state that a sample should be as large as the researcher can obtain within the given time and energy. The larger the sample, the more accurate the results as it would be more representative of all subgroup types in the overall population. With these arguments in mind, the researcher decided on a sample size of fifty four students, three scorers and five lecturers.

In this study, the judgment or purposive sampling procedure was used to make the elicitation of data more economical, efficient and practical. Babbie (1999) asserts that it is appropriate for researchers to select their own sample based on prior knowledge of the population, its elements and the aim of the research. Therefore, the subjects were selected for inclusion in the sample, based on the judgement of the researcher. As mentioned by Babbie (1999), there is the tendency for sampling bias to occur with this method of sampling whereby the subjects selected for the sample are not typical or representative. In order to reduce this element of sampling bias, the subjects were required to fulfill several criteria.

3.3.1 Student Selection Criteria

The researcher identified fifty-four students from a pre- university student population of 2500 students in TCSJ. Each of the students had to be enrolled in

any one of the pre-university programmes ('A' levels, South Australian Matriculation or Canadian Pre University) at TCSJ. They had to also be enrolled in a 16-hour intensive IELTS preparation course with the English Language Centre of TCSJ.

It was also necessary for the subjects to have obtained a minimum grade of P3 in the SPM examinations. This was to ensure that only students with an upper-intermediate English Language proficiency level would be participating in this study. The SPM examination was used as a yardstick to measure students' proficiency because this would mean that the type of language training and general education the student would have been exposed to and the level of language proficiency would be similar.

Furthermore, it was compulsory for all students participating in this research to attend all classes and complete the given homework. Participants, who missed a class or failed to complete any one of the given assignments, were excluded from the sample. The same lecturer (the researcher) taught both the courses (Course A and Course H) from which the students were sampled in order to ensure the consistency in teaching methods and materials received by every sample. With the fulfillment of the criteria, the input received by each student throughout the duration of the course would be the same. Since the samples were students of the researcher, the above-mentioned criteria were strictly adhered to.

3.3.2 Scorer Selection Criteria

Three scorers were identified to score the students' essays. This was to increase the reliability of the scores. Weir (1990) asserts, "...essays marked independently by more than one marker with marks averaged is a more reliable estimate than if it were marked by a single marker" (Weir 1990, p.165). However, Weir (1990) also cautions that this would only be possible if there exist an adequate marking scheme and if there is some form of standardization amongst the scorers. Hence, the scorers were required to fulfill two criteria.

The first criterion required all scorers to be trained in ESL and have a minimum of three years teaching experience in ESL. This criterion was imposed in order to ensure that the scorers were familiar with learning and teaching strategies in the ESL classroom and had prior experience in grading essays. The scorers were also required to be familiar with the IELTS examination format and marking scheme. The fulfilment of the above mentioned criteria will enable the scorers to score the students essays in a more valid and reliable manner.

3.3.3 Selection Criteria for Interviewees

Five lecturers of TCSJ were sampled for the purposes of this study. The lecturers were ESL trained, with a minimum of three years teaching experience in the ESL field. In addition, it was also necessary for the lecturers to have used and be

familiar with a variety of scoring methods – namely, the holistic and analytic methods of scoring essays. This was necessary in order for them to be able to give opinions on the methods of scoring essays. Finally, they were required to be involved in the teaching of intensive preparatory courses, as this would ensure that they were well aware of the demands of such a course.

3.4 Instruments

The instrumentation section of this chapter describes the measures that were employed in this study and how they were used to measure the variables specified in the research questions. There is no one perfect method to use to conduct all surveys as all the different methods have their pros and cons. According to Frankeal and Wallen (2000), several criteria such as cost, implementation time, response rates, available technology, availability of sample etc. should be considered before making decisions about which instrument to choose for the data collection process. Information can be obtained by direct measurement, observation or interviews. For the purposes of this research, three instruments were used to collect data - essays, questionnaires and interviews.

3.4.1 Students' Essays (Pre-tests & Post-tests)

A pre-test and a post-test were used to collect the students' essays necessary for this study. The main aim of these two tests was to identify if there exists any

relationship(s) between the methods of scoring used and students' writing performance before and after they were exposed to the different scoring methods. Students were required to write on the same topic in both tests. The only difference was the time of administration. The pre-test was administered at the beginning of the course (an hour before the first lesson commenced) and the post-test was conducted at the end of the course (the last hour upon concluding the last lesson). Both tests consisted of one similar essay question (Task 2). The test question was taken from a popular IELTS practice test book 'IELTS Practice Now'. The students were instructed to write an academic essay for a non-specialist audience on the topic "Zoos are sometimes seen as necessary but poor alternatives to a natural environment. Discuss some of the arguments for and against keeping animals in zoos." A duration of 40 minutes was allocated for the completion of this task and the essay should be about 250 words in length. The specifications for this task followed those of the IELTS Writing Task 2 test specifications.

3.4.2 Students' Questionnaires

One commonly used method of data collection in survey research is the questionnaire. Questionnaires can provide us with a wealth of descriptive data pertaining to individuals or groups. The questionnaire is used as a tool to collect data as it can cover a large sample in a short span of time for gathering highly structured data. It is also a rather inexpensive way of collecting data. In this study,

the questionnaire was administered to all 54 participants. The main aim of this tool was to elicit information about students' opinions on the two methods of scoring.

The main concern in the construction of the questionnaire is reliability and validity. This study employed a modified version of a questionnaire developed by Ferris (1995) to obtain feedback on each student's compositions. The questionnaire consisted of seven questions; six closed questions and one open question. The first question asked students to rate their ability in composition writing. The aim of this question was to give the researcher some background knowledge about how the students perceived their writing skills. The second question was concerned about the student's response when they received their essays with just a grade placed at the end of the paper. In the third question, they were then asked if they expected their teachers to correct all the mistakes they made in their essays. The fourth question attempted to find out whether the students made any effort to re-read their essays once their lecturer returned it to them. Question five attempted to seek whether the students undertook any effort to think about their mistakes and self-edit their work. In question six, the students were required to state the pro-active action(s) they would take after reading their teacher's comments and corrections. Question seven expected students to give their opinions on whether their teacher's responses and comments so far have helped improve their composition writing skills (Refer to Appendix A).

3.4.3. Interviews with Lecturers

An interview is usually the choice where fewer subjects are involved and when the required data relate to individual experience and the evaluator wishes to explore in some depth the opinions, expectations and actions of individuals. The primary concern of these interviews was to get information regarding the method of scoring that the lecturers used, their personal opinions about these methods of scoring and the responses they have received from their students in the course of their teaching experience. This instrument (the interview question guide) was developed by the researcher personally as a standard instrument was not available.

Five TCSJ lecturers with an ESL background were interviewed. The interview was a semi-structured one and the questions were divided into five sections. The first section contained three questions aimed at eliciting information about the lecturers' careers and educational background. The second section contained two questions about the teachers' viewpoints on the methods of scoring essays in intensive language proficiency courses. The following section covered four questions on analytic scoring, its popularity, how students respond to this method of scoring, whether students read all the teachers' comments and whether it helped them in improving their writing performance. The fourth section comprised three questions on holistic scoring. The first question was about its popularity, the next question asked about how students responded to this method

of scoring and the third if this method of scoring would aid students to improve their writing performance. The fifth section contained three questions directed to the teacher on other methods of scoring compositions. The first question asked lecturers to suggest alternative methods of scoring, the next question enquired if they used these alternative methods to score compositions, and the last question was about whether these alternative scoring methods are more effective in improving student writing (Refer to Appendix B).

3.5 Scoring Scheme

“Scoring schemes are developed by teachers or other evaluators to guide the analysis of the products or processes of students’ efforts,” says Moskal (2000, p.1). One common use of scoring schemes is to guide the evaluation of writing samples. In order for the scorers to grade the essays written by the subjects of this study, two scoring schemes were chosen - the IELTS exam band descriptor scale developed by the ETS (English Testing Services) and the ESL Composition Profile Analytic Scale developed by Jacobs et.al (1981). These two schemes are rather popular scoring schemes and are used widely in ESL classrooms. According to Clarke (1999), using popular or previously used scoring scales is an advantage, as they would have been subjected to testing in terms of validity and reliability.

3.5.1 The IELTS Scoring Scale

Table 3.1 below shows the IELTS Scoring Scale which is a holistic scoring system made up of a band descriptor scale. There are ten bands in total from 0 to 9. Each band has a short phrase such as 'Non User', 'Limited User', 'Good User' etc. which corresponds to a descriptive statement giving a summary of the English proficiency of a candidate classified at that level. If a student does not attempt the question at all, only then is he or she awarded a band of 0. In order to obtain a band of 9 (the highest band), a student has to exhibit a near native command of the language.

Table 3.1: The IELTS Band Scoring Scale

9	EXPERT USER	Has fully operational command of the language: appropriate, accurate and fluent with complete understanding.
8	VERY GOOD USER	Has fully operational command of the language with only occasional unsystematic inaccuracies and inappropriacies. Misunderstandings may occur in unfamiliar situations. Handles complex detailed argumentation well.
7	GOOD USER	Has operational command of the language, though with occasional inaccuracies, inappropriacies and misunderstandings in some situations. Generally handles complex language well and understands detailed reasoning.
6	COMPETENT USER	Has generally effective command of the language, despite some inaccuracies, inappropriacies and misunderstandings. Can use and understand fairly complex language, particularly in familiar situations.
5	MODEST USER	Has partial command of the language, coping with overall meaning in most situations, though is likely to make many mistakes. Should be able to handle basic communication in own field.
4	LIMITED USER	Basic competence is limited to familiar situations. Has a frequent problem in understanding and expression. Is not able to use complex language.
3	EXTREMELY LIMITED USER	Conveys and understands only general meaning in very familiar situations. Frequent breakdowns in communication occur.
2	INTERMITTENT USER	No real communication is possible except for the most basic information using isolated words or short formulae in familiar situations and to meet immediate needs. Has great difficulty understanding spoken and written English.
1	NON USER	Essentially has no ability to use the language beyond possibly a few isolated words.
0	DID NOT ATTEMPT THE TEST	No assessable information provided.

3.5.2 The ESL Composition Profile Scale

Table 3.2 shows the ESL Composition Profile Scale is an analytic scoring scale. It gives weight to five components of writing – content (30%), organization (20%), vocabulary (20%), language use (25%) and mechanics (5%). Within each component, the scores are broken down into several ranges and each range is represented by a set of criteria. For example, if a student obtains a score of 22 out of 30 for content, he or she will fall under the 'Fair to Poor' range. On the other hand, a score of 27 out of 30 would put the student in the 'Excellent to Very Good' range.

Table 3.2: The ESL Composition Profile Scale

COMPONENT	SCORE	CRITERIA
C O N T E N T	30-27	EXCELLENT TO VERY GOOD Knowledgeable Substantive development of thesis Relevant to assigned topic
	26-22	GOOD TO AVERAGE Sure knowledge of subject Adequate range Limited development of thesis Mostly relevant to topic but lacks detail
	21-17	FAIR TO POOR Limited knowledge of subject Little substance Inadequate development of topic
	16-13	VERY POOR Does not show knowledge of subject Non-substantive Not pertinent OR not enough to evaluate

O R G A N I Z A T I O N	20-18	EXCELLENT TO VERY GOOD Fluent expression Ideas clearly stated/supported Succinct Well-organized Logical sequencing Cohesive
	17-14	GOOD TO AVERAGE Somewhat choppy Loosely organized, main ideas stand out Limited support Logical but incomplete sequencing
	13-10	FAIR TO POOR Non-fluent Ideas confused or disconnected Lacks logical sequencing
	9-7	VERY POOR Does not communicate No organization OR not enough to evaluate
V O C A B U L A R Y	20-18	EXCELLENT TO VERY GOOD Sophisticated range Effective word/idiom choice and usage Word form mastery Appropriate register
	17-14	GOOD TO AVERAGE Adequate range Occasional errors of word/idiom etc. Meaning not obscured
	13-10	FAIR TO POOR Limited range Frequent errors of word/idiom form, choice, usage Meaning confused or obscured
	9-7	VERY POOR Essentially translation Little knowledge of English vocabulary, idioms, word form OR not enough to evaluate

L A N G U A G E U S E	25-22	EXCELLENT TO VERY GOOD Effective, complex constructions Few errors of agreement, tense, number, word order/function, articles, pronouns, prepositions
	21-18	GOOD TO AVERAGE Effective but simple constructions Minor problems in complex constructions Several errors of agreement, tense, number, word order/function, articles, pronouns, prepositions, but meaning seldom obscured
	17-11	FAIR TO POOR Major problems in simple/complex constructions Frequent errors of negation, agreement, tense, number, word order/function, articles, pronouns, prepositions and/or fragments, run-ons, deletions Meaning confused or obscured
	10-5	VERY POOR Virtually no mastery of sentence construction rules Dominated by errors Does not communicate OR not enough to evaluate
M E C H A N I C S	5	EXCELLENT TO VERY GOOD Demonstrates mastery of conventions Few errors of spelling, punctuation, capitalization, paragraphing
	4	GOOD TO AVERAGE Occasional errors of spelling, punctuation, capitalization, paragraphing, but meaning not obscured
	3	FAIR TO POOR Frequent errors of spelling, punctuation, capitalization, paragraphing Poor handwriting Meaning confused or obscured

	2	VERY POOR No mastery of conventions Dominated by errors of spelling, punctuation, capitalization, paragraphing Handwriting illegible OR not enough to evaluate
--	---	--

3.6 Methodology

The data collection procedure was carried out over a period of two months. The entire procedure was administered in five stages:

1. Selection of Subjects
2. Training of the Scorers
3. Administration of Pre –test and Post-test (Course A & H)
4. Administration of Questionnaires
5. Interviews with the lecturers

3.6.1 Subjects, Writing Tasks and Scoring Methods

The first stage undertaken was the selection of subjects. The 54 students sampled for this study were divided into two groups, Group A and Group H. They were then placed into Course A and Course H respectively. Both groups of students were given sixteen hours of classroom instruction. During this period, the four skills (listening, speaking, reading and writing) were taught as required by the course and six hours were devoted to the teaching of the writing component.

Task 2 of the writing component was used for the investigation. The students in both groups were exposed to three essay types; discursive, cause-effect and argumentative. They were given six Task 2 writing practices over the duration of four weeks. The teaching materials and questions used in both courses were the same. The only difference between the two groups was that students in Course A were exposed to the analytic scoring method throughout the course. Hence, the lecturer made specific comments and scored each of their essays using the analytic scoring technique (ESL Composition Profile Scoring Scale). Students in course H however, were exposed only to the holistic method of scoring. Each of the essays was scored holistically. The lecturer just gave them a band score at the end of each piece of writing.

3.6.2 Training of Scorers

The second stage undertaken was the training of the scorers. Fraenkel & Wallen (2000) advocate that scoring agreement required a demonstration that independent scorers can achieve satisfactory agreement in their scoring. Therefore, the researcher requested an ex-teacher from TCSJ, who is currently a licensed IELTS examiner with IDP Education Australia located in Kuala Lumpur to read through about 15 sample essays and score them in accordance to the IELTS bands (0-9). These scripts were then used as benchmark papers. Copies were made and distributed to the scorers participating in this research. Along with this, the scorers

were also given a copy of the two scoring scales and were told to use the analytic scale to score the benchmark papers. The scorers were given a few days to do this.

Next, a standardization meeting was held with the three scorers. The scorers had the opportunity to clarify any doubts on the use of the two scoring schemes. This procedure for training the scorers was carried out in line with the guidelines advocated by Fraenkel & Wallen (2000) and Reid's (1993) suggestion that "...with the assistance of scoring guide that supplies general criteria for consideration, with training of readers by experienced raters and with benchmark papers, a high rate of consensus between two scorers can be achieved" (Reid 1993:238).

3.6.3 Administration of Pre-test and Post-test

The next step involved the administering of the pre-test for the students in Course A and Course H. The researcher herself conducted the test as she was also the lecturer for both courses (Course A and Course H). It was held on the first lesson of the course. The students were given 40 minutes to complete the specified writing task.

The pre-test papers were then collected and handed to the scorers. The scorers were given three weeks to score the papers. The scorers had to give each essay two separate scores; one score based on the IELTS Band Scale and another based

on the Composition Profile analytic scale. Once the scoring was completed, the papers were returned to the researcher.

The post-test was administered on the last lesson of the respective courses and the same essay topic was given. Upon completion of the task, the papers were once again given to the scorers to be graded, following the two scoring methods used to grade the pre-tests. After this, they were returned to the researcher.

3.6.4 Administration of the Questionnaires

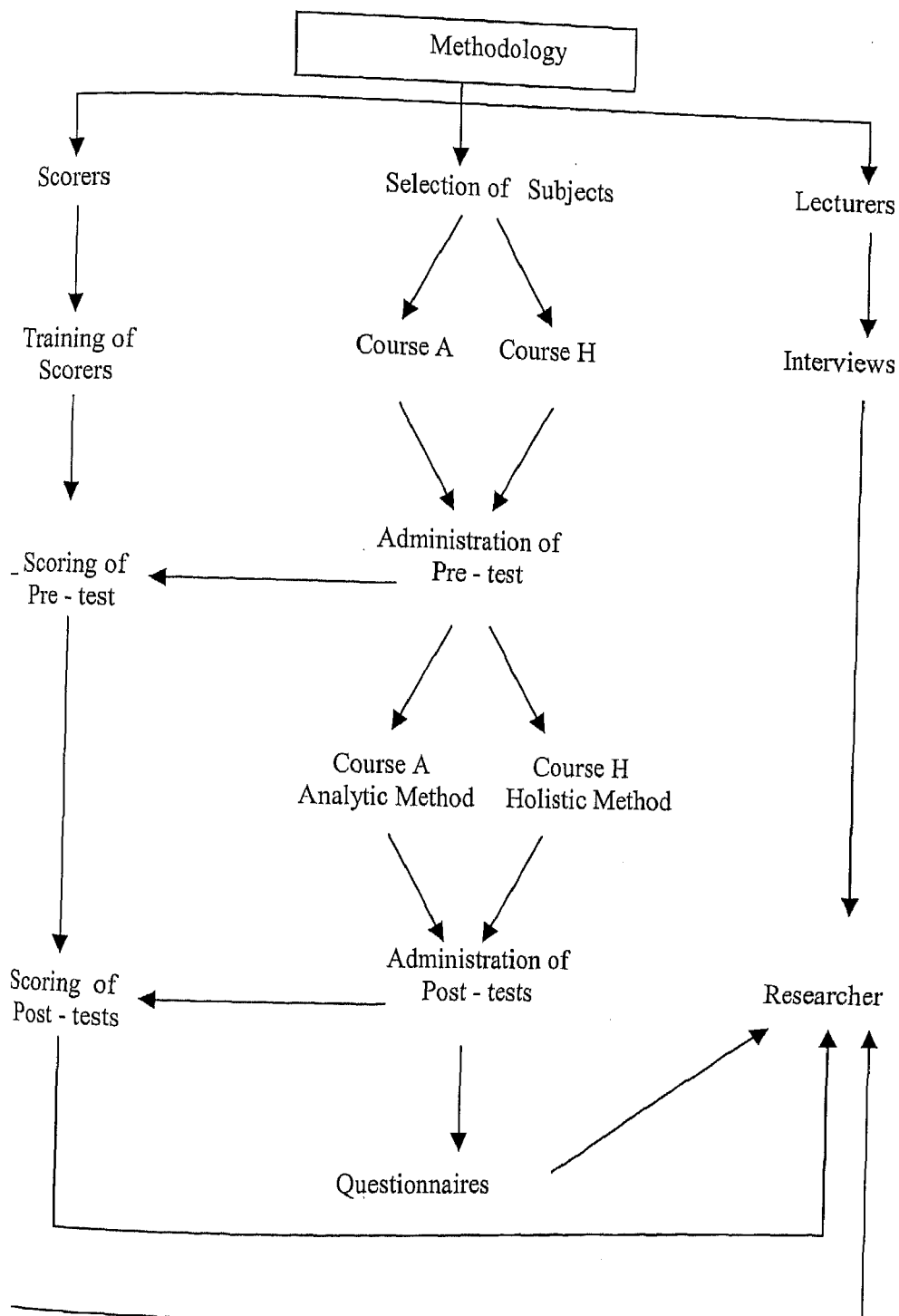
The administration of the questionnaire surveys were carried out on the last day of the course. The students in Course A and Course H were each given a questionnaire to complete. They were told to answer each question and were reminded to take it seriously as their opinions mattered very much. This task took approximately 20 minutes and was conducted by the researcher.

3.6.5 Interviews with Lecturers

Finally, the five language teachers who were identified earlier were interviewed.

The researcher conducted these interviews in an informal setting at the college premises. It took approximately three weeks to complete all five interviews. Figure 3.1 shows the four stages (which were earlier discussed) involved in the data collection process.

Figure 3.1: The Various Stages of the Data Collection Process



3.7 Data Analysis

Upon completion of the administration of the various instruments, the collation and tabulation of data was undertaken. The essay scores from the three scorers for both the pre-test and post-test were averaged and presented in tables.

The responses from the questionnaires were analysed and tabulated. The raw scores were then converted into percentages. Since the last question elicited qualitative data, this data was summarized according to the variety of responses. The various stages of data collection are presented in figure 3.1.

The responses from the interviews were qualitative in nature hence were summarized in accordance to the five major themes in the interview guide - Teacher's Background, Teacher's Viewpoint on Methods of Scoring Essays, Analytic Scoring, Holistic Scoring and Other Methods of Scoring.

This chapter has described and explained the research design and procedures that were adopted in detail. The following chapter will now present the findings of this research.